

控制与决策

Control and Decision

离异图消歧引导的偏标记学习

胡峰, 刘鑫, 邓维斌, 代劲, 刘群

引用本文:

胡峰, 刘鑫, 邓维斌, 代劲, 刘群. 离异图消歧引导的偏标记学习[J]. *控制与决策*, 2023, 38(6): 1753–1760.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.1845>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度信念网络和迁移学习的隐匿FDI攻击入侵检测

Stealthy FDI attack detection based on deep belief network and transfer learning

控制与决策. 2022, 37(4): 913–921 <https://doi.org/10.13195/j.kzyjc.2020.1469>

基于多标签学习的旋转机械分级复合故障诊断

Hierarchical compound fault diagnosis of rotating machinery based on multi-label learning

控制与决策. 2022, 37(7): 1772–1778 <https://doi.org/10.13195/j.kzyjc.2021.0067>

基于自适应多尺度图卷积网络的多标签图像识别

Multi-label image recognition based on adaptive multi-scale graph convolutional network

控制与决策. 2022, 37(7): 1737–1744 <https://doi.org/10.13195/j.kzyjc.2021.0179>

实体消歧综述

Entity disambiguation: A review

控制与决策. 2021, 36(5): 1025–1039 <https://doi.org/10.13195/j.kzyjc.2020.0388>

基于共享隐空间的多视角SVM

Multi view SVM based on common hidden space

控制与决策. 2021, 36(3): 534–542 <https://doi.org/10.13195/j.kzyjc.2019.0829>

离异图消歧引导的偏标记学习

胡峰^{1,2†}, 刘鑫¹, 邓维斌², 代劲², 刘群²

(1. 重庆邮电大学 计算机学院, 重庆 400065; 2. 重庆邮电大学 计算智能重庆市重点实验室, 重庆 400065)

摘要: 偏标记学习是一种弱监督学习框架, 它试图从样本的多个候选标签中选择唯一正确的标签. 消歧是偏标记学习中的一种重要手段, 主要通过算法判别潜在的真实标签. 目前, 人们普遍采用单一的特征空间或者标签空间进行消歧, 容易导致算法受到不准确先验知识的引导而陷入鞍点. 针对消歧过程中特征相似样本易受到异类样本影响从而影响消歧效果这一问题, 定义了样本离异点和离异图; 在此基础上, 提出一种离异图引导消歧的偏标记学习方法. 该方法利用标签空间的差异构建离异图, 可以有效结合特征空间的相似性和标签空间的差异性, 降低离异点为消歧过程带来的潜在风险. 实验结果表明, 与 PLKNN、IPAL、SURE、PL-AGGD、SDIM、PL-BLC、PRODEN 等方法相比较, 所提出的算法在偏标签学习方法中表现更好, 能够取得良好的消歧效果.

关键词: 弱监督; 偏标记学习; 消歧; 流形假设; 最大间隔; 图嵌入

中图分类号: TP181

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1845

引用格式: 胡峰, 刘鑫, 邓维斌, 等. 离异图消歧引导的偏标记学习 [J]. 控制与决策, 2023, 38(6): 1753-1760.

Introducing outlier graph to disambiguate for partial label learning

HU Feng^{1,2†}, LIU Xin¹, DENG Wei-bin², DAI Jing², LIU Qun²

(1. College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Partial label learning is a weakly supervised learning framework, which attempts to select the only correct label from multiple candidate labels of the sample. Disambiguation is an important means in partial tag learning, which mainly uses algorithms to identify potential real tags. At present, researchers generally use a single feature space or label space for disambiguation, which is easy to lead the algorithm to fall into saddle point under the guidance of inaccurate prior knowledge. In order to solve the problem that similar samples are easy to be affected by abnormal samples in the process of disambiguation, sample outlier and outlier graph are defined. On this basis, a partial marker learning method for dissociation graph guided disambiguation is proposed. The algorithm uses the difference of label space to construct the divorced point map, which can effectively combine the prior knowledge of feature space and label space, and restrict each other to resist the potential high risk brought by divorced samples to the disambiguation process. Experimental results show that compared with PLKNN, IPAL, SURE, PL-AGGD, SDIM, PL-BLC and PRODEN, the proposed algorithm performs better in the partial label learning method and achieves good disambiguation effect.

Keywords: weak supervision; partial label learning; disambiguation; manifold hypothesis; maximum interval; graph embedding

0 引言

传统的多类分类问题属于有监督学习, 基于强监督假设建模, 其样本的类别标记信息是唯一明确的. 偏标记学习 (partial label learning) 属于弱监督学习, 也被称为模糊标记学习 (ambiguous label learning)^[1-3] 和超集标记学习 (superset label learning)^[4-6],

其目的是学得一个多类分类器. 与多分类问题不同的是, 它的每个训练样本被分配一组候选标签, 其中只有一个是有效的.

现实世界中往往很难获得准确标记信息, 因此, 偏标记学习已成功应用于各个实际应用领域中. 例如, 在众包在线注释系统中 (图 1(a))^[7], 不同的

收稿日期: 2021-10-27; 录用日期: 2022-03-15.

基金项目: 国家重点研发计划项目 (2018YFC0832102); 重庆市高校与中国科学院所属院所重点合作项目 (HZ2021008); 重庆市自然科学基金项目 (cstc2021jcyj-msxmX0849).

†通讯作者. E-mail: hufeng@cqupt.edu.cn.

用户可能会采用不同的标签标注同一幅图片,而真实的标签信息则隐藏在用户所标注的候选标记集中. 另一个比较具有代表性的应用是自动人脸标注(图1(b))^[3],即使用文本标题中包含的名称集合来对图像中出现的人物面孔进行命名. 此外,偏标记同样也成功地应用于生态信息学^[4]、自然语言处理^[8]等领域.



用户A的注释:狸花猫用户
用户B的注释:猓猓
用户C的注释:兔猓

(a) 众包在线注释



BBC news: 美国大选, 从特朗普到拜登, 总统换届过度交接如何进行

(b) 自动人脸标注

图1 偏标记应用示例

在偏标记学习中,由于处理的标记信息不满足单一性和明确性,而是一个候选标记集合的形式,这使得学习建模变得十分困难. 由于训练集中的标签信息是模糊的,如何消除候选标签的歧义是偏标记学习任务的关键. 现有的偏标记学习算法主要采用两种不同的消歧策略:基于平均的消歧和基于辨识的消歧.

基于平均的消歧方式对所有候选标签一视同仁,比较有代表性的算法包括基于样本特性的消歧算法和基于模型参数的消歧算法. 基于样本特性的消歧算法,如PL-KNN^[1]、RegISL^[6],它们通过对训练样本的邻居求平均值来预测测试样本的标签. 基于模型参数的消歧算法,如CLPL^[2],该方法通过区分候选标签和非候选标签上的平均建模输出来达到消歧目的. 由于在训练过程中没有实现标签识别,基于平均的消歧算法直观简洁,便于实现和理解,但是,该类算法忽略了候选标签之间的差异,真实标签容易被大量假正标签误导.

基于辨识的消歧方式将偏标记对象的真实标记作为隐变量,通过EM算法(expectation-maximization

algorithm)流程迭代优化内嵌隐变量的目标函数实现消歧. Zhang等^[9]提出了PL-LEAF方法,将学习过程分为消歧和预测两个阶段:在消歧阶段采用KNN找到在特征空间上相似的样本,并在特征空间上构建样本权重图;在此基础上,结合平滑假设,将特征空间权重推广到标签空间,利用样本特征相似度进行消歧并取得了较好的效果. 后有学者发现,PL-LEAF在平滑性假设中容易受到样本离异点影响,使得从原始特征空间恢复的流形结构可靠性降低,据此提出了一种基于自适应图的优化方法:PL-AGGD算法^[10],该方法结合消歧和相似图权重更新进行联合优化,在实验中取得了更好的结果. Feng等^[11]利用标记空间中的潜在有用信息进行消歧,提出了语义差异最大化(SDIM)方法,并证明了通过标签空间信息进行消歧的可行性. 结合标签空间和特征空间的差异性,陈鸿昶等^[12]提出了CAP算法(candidate label-aware partial label learning, CLAPLL, 简称CAP),该方法通过构建示例特征相似图和示例标记集相似图来度量偏标记数据间的相似度,在一定程度上减少了离异点对相似图构造的影响,但这种方法会带来额外的误差传递,且算法不能在训练中动态更新图信息,导致其仍然不能摆脱样本离异点的影响. Yan等^[13]通过批量标签校正做数据增强来提高性能. Lv等^[14]、Cabannes等^[15]、Wen等^[16]则相继在基于辨识的模型上对优化项和损失函数进行调整,也取得了较好的效果.

基于辨识的消歧策略,不足之处在于容易受到不准确的先验知识引导而导致模型过拟合. 本文提出一种新的消歧方法PL-ODG (introducing outlier graph to disambiguate for partial label learning),该方法利用标签空间的语义差异化信息,弥补相似图引导消歧的不足,构建更为稳定的图嵌入算法来引导消歧. 首先,结合KNN算法构建初步的特征相似图;其次,在相似图中去除离异点,构建新的相似图,并利用标签空间的差异构建离异图;然后,结合相似图和离异图进行消歧;最后,通过大量的现实世界数据集的结果表明,本文所提出的方法体现了出色的消歧性能.

本文的创新在于定义了样本离异点和离异图,并利用离异点和离异图作为启发式进行消歧. 该方法在消歧过程中,同时嵌入了特征空间和标记空间的先验知识,提出了一种更为稳定的消歧算法.

1 离异图引导的消歧算法

1.1 相似图和离异图

文中需特别说明的内容主要有以下几个方面:文献[9]和文献[10]提出了特征相似图引导消歧的偏

标记学习方法 PL-LEAF 和 PL-AGGD, 这两种方法皆采用 KNN 判定样本是否相似, 并提前设定邻居个数 K . 本文通过对样本的分析发现, 偏标记学习中存在这样的现象: 部分样本特征空间相似而标签空间差异较大; 在偏标记学习中, 样本的真实标记信息隐藏在候选标记集合中, 即候选标记集合没有交集的样本一定不属于同一类别. 此外, 在特征空间相似而候选标记集合不重合的样本对于指导消歧具有重要意义(本文称为样本离异点), 如果对这些样本离异点进行有针对性的处理, 则可能会对消歧产生积极的作用. 下面给出相关概念的定义.

令 $X \in \mathbf{R}^d$ 表示样本空间, $Y = [y_1, y_2, \dots, y_m]^T \in \{0, 1\}^{m \times q}$ 表示标记空间, 其中 $y_{ij} = 1$ 表示第 i 个样本的第 j 个候选标签为正.

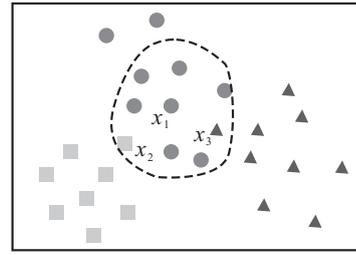
定义 1 (特征相似样本) 给定样本 $x_i \in X$, $N(x_i) = \{x_j | x_j \in \text{KNN}(x_i), i \neq j\}$ 表示由 KNN 算法所产生的 x_i 的 K 个邻居的样本集合; $\forall x_j \in N(x_i)$, 若 x_i 与 x_j 在特征空间上相似, 则称 x_j 为 x_i 的特征相似样本.

定义 2 (样本离异点) 若 x_j 为 x_i 的特征相似样本, 且它们的标签候选集交集为空, 即 $x_j \in \text{KNN}(x_i) \cap y_i y_j^T = 0$, 则称 x_j 为 x_i 的样本离异点.

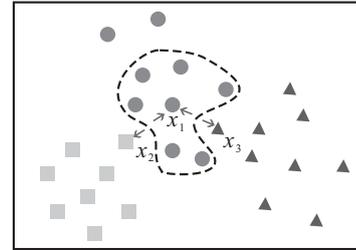
定义 3 (相似图) 给定一个加权图 $G_1 = (V, \xi_1, S_1)$. 其中: 顶点集 $V = \{x_i | 0 \leq i \leq m\}$ 表示所有训练样本. 边集 $\xi_1 = \{(x_i, x_j) | x_i \in \text{KNN}(x_j) \cap y_i y_j^T \neq 0, i \neq j\}$ 表示顶点之间的关系, 且满足 $\forall x_i \in V$, 若 $x_j (x_j \in N(x_i))$ 不是 x_i 的样本离异点, 则 $(x_i, x_j) \in \xi_1$. $S_1 \in \mathbf{R}^{m \times m}$ 表示权重矩阵, 当 $(x_i, x_j) \notin \xi_1$ 时, $s_{ij} = 0$; 当 $(x_i, x_j) \in \xi_1$ 时, $s_{ij} = 1$. 则称加权图 $G_1 = (V, \xi_1, S_1)$ 为相似图.

定义 4 (离异图) 给定一个加权图 $G_2 = (V, \xi_2, S_2)$. 其中: 顶点集 $V = \{x_i | 0 \leq i \leq m\}$ 表示所有训练样本. 边集 $\xi_2 = \{(x_i, x_j) | x_i \in \text{KNN}(x_j) \cap y_i y_j^T = 0, i \neq j\}$ 表示顶点之间的关系, 且满足 $\forall x_i \in V$, 若 $x_j (x_j \in N(x_i))$ 是 x_i 的样本离异点, 则 $(x_i, x_j) \in \xi_2$. $S_2 \in \mathbf{R}^{m \times m}$ 表示权重矩阵, 当 $(x_i, x_j) \notin \xi_2$ 时, $s_{ij} = 0$; 当 $(x_i, x_j) \in \xi_2$ 时, $s_{ij} = 1$. 则称加权图 $G_2 = (V, \xi_2, S_2)$ 为离异图.

样本离异点在指导消歧的过程中具有重要作用, 这些点往往就是样本中难以区分的困难样本, 即容易被误分类成相邻类别的样本. 若只利用特征相似度进行消歧, 则在特征空间上相似的样本将会被先验知识所引导, 从而被误判为相同类别. 下面给出一个示例进行说明. 图 2(a) 中的 x_2 和 x_3 是 x_1 的样本离异点 (KNN 选取近邻时 $K = 8$). 如果只考虑特征的相似性, 则 x_2 和 x_3 将会被标记为 x_1 的同样类别.



(a) 圆形的特征相似空间



(b) 修正后的特征相似空间

图 2 样本离异点对特征空间流形假设的影响 (图中不同形状表示不同的真实类别)

针对这类样本, 可以利用最大样本间隔思想, 使得训练过程中保持样本离异点之间距离最大. 即: 一方面满足中心点与特征相似样本间的距离尽可能小; 另一方面, 满足中心点与样本离异点间的距离尽可能大. 通过这种处理, 可以有效引导样本消歧, 从而得到稳定的最优解. 整个优化过程将会结合凹函数和凸函数, 后文将给出具体求解方法. 在示例图 2(a) 中, 经过上述处理, 可以得到如图 2(b) 所示的效果.

在相似图 G_1 和离异图 G_2 中, 令 S_1 和 S_2 分别是对应的非负权重矩阵, 根据这两个矩阵, 可以采用迭代学习的策略. 对于 S_1 , 可通过最小化相似样本之间的距离进行迭代更新, 以达到自适应相似度的效果, 具体迭代过程可参考下式:

$$\begin{aligned} \min_{S_1} \sum_{j=1}^m \left\| x_j - \sum_{(x_i, x_j) \in \xi_1} s_{ij} x_i \right\|_2^2; \\ \text{s.t. } 1 \leq i \leq m, \sum_i s_{1ji} = 1. \end{aligned} \quad (1)$$

对于 S_2 , 则需要进行不同的处理. 由于离异图中的样本具有特征空间相似而标签不同的特性, 可用 0 和 1 表示样本间是否邻接且标签相异, 并进行归一化约束即可, 无需对其权重再进行更新.

1.2 消歧算法

设真实候选标签集合为 $Y = [y_1, y_2, \dots, y_m]^T \in \{0, 1\}^{m \times q}$, 置信度矩阵为 $P = [p_1, p_2, \dots, p_m]^T, 0 \leq p \leq 1$.

根据平滑性假设, 特征空间的流形结构应该保存在标记空间中. 根据最大间隔理论, 不同类别样本的最优划分方式为不同类别样本之间的距离最大

化. 结合平滑性假设和最大间隔理论, 可以利用相似图 G_1 和离异图 G_2 得到如下两个公式:

$$\min_P \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_1} s_{1ij} p_i \right\|_2^2, \quad (2)$$

$$\max_P \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_2} s_{2ij} p_i \right\|_2^2. \quad (3)$$

式(2)和(3)同时满足 $\forall 1 \leq i \leq m, 0 \leq P \leq Y, \sum_i P_{ji} = 1, \sum_i s_{1ji} = 1, \forall s_{2ji} > 0, \sum_i s_{2ji} = 1$.

式(2)是凸函数, 而式(3)是凹函数. 为了便于后面优化, 组合式(2)和(3)可以得到

$$\begin{aligned} & \min_P \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_1} s_{1ij} p_i \right\|_2^2 - \\ & \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_2} s_{2ij} p_i \right\|_2^2. \\ & \text{s.t. } \forall 1 \leq i \leq m, 0 \leq P \leq Y, \\ & \sum_i P_{ji} = 1, \sum_i s_{1ji} = 1; \\ & \forall s_{2ji} > 0, \sum_i s_{2ji} = 1. \end{aligned} \quad (4)$$

通过式(4)可以得到一个置信度矩阵 P , 即消歧结果, 可作为新的标签用于模型训练. 接下来, 利用带 L_2 正则的多元回归模型拟合训练得到的置信度矩阵 P . 令 $L(x, W)$ 表示模型输出, 用最小二乘法求解, 同时使用 L_2 正则来减少过拟合的风险, 用超参数 λ 来控制其权重, 有

$$\min_W \|L(x, W) - P\|_2^2 + \lambda \|W\|_2^2. \quad (5)$$

为了得到更好的效果, 将式(1)、(4)、(5)组合进行求解, 以便在模型的迭代更新过程中对 P 和 S_1 进行动态优化, 减少先验误差. 在这里额外加入了3个超参数 μ, η, γ , 分别用于调整最小化相似样本距离、最大化离异样本距离、更新 S_1 的权重, 以进一步增加模型泛化能力, 后文将通过实验给出超参数的优化策略. 最终得到的算法公式如下:

$$\begin{aligned} & \min_{S_1, P, W} \sum_{i=1}^m \|L(x_i, W) - P\|_2^2 + \\ & \mu \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_1} s_{1ij} p_i \right\|_2^2 - \\ & \eta \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_2} s_{2ij} p_i \right\|_2^2 + \\ & \gamma \sum_{j=1}^m \left\| x_j - \sum_{(x_i, x_j) \in \xi_1} s_{1ij} x_i \right\|_2^2 + \lambda \|W\|_2^2. \\ & \text{s.t. } \forall 1 \leq i \leq m, 0 \leq P \leq Y, \end{aligned}$$

$$\begin{aligned} & \sum_i P_{ji} = 1, \sum_i s_{1ji} = 1; \\ & \forall s_{2ji} > 0, \sum_i s_{2ji} = 1. \end{aligned} \quad (6)$$

考虑到需要优化3个变量: W, P 和 S_1 , 同时优化的难度过大, 本文将采用EM算法的思想, 通过固定多余变量的方式进行交替求解, 即每次固定其余的变量, 只迭代优化一个变量.

2 优化方法

初始化 P 矩阵权重时, 一般采用两种策略: 1) 将候选标记集合一视同仁都置为1, 进行一次迭代后再进行归一化; 2) 直接将 Y 进行归一化作为初始的 P 矩阵权重, 即 $\forall 1 \leq i \leq m \left(\sum_j P_{ij} = 1 \right)$. 对于第2种方法, 可在第1次学习的过程中, 对歧异度小的样本赋予更高关注度, 它更适用于样本总体歧异度偏小的情况. 但当样本总体歧异度偏大时, 真实标签更容易淹没在候选标记集中, 会导致大量样本被模型忽视.

在得到初始 P 权重后, 接下来将依次按下面步骤进行迭代优化, 直到算法收敛.

首先, 固定 P 和 W , 优化 S_1 . 当固定 P 和 W 后, 有

$$\begin{aligned} & \min_{S_1} \mu \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_1} s_{1ij} p_i \right\|_2^2 + \\ & \gamma \sum_{j=1}^m \left\| x_j - \sum_{(x_i, x_j) \in \xi_1} s_{1ij} x_i \right\|_2^2; \\ & \text{s.t. } 1 \leq i \leq m, \sum_i s_{1ji} = 1. \end{aligned} \quad (7)$$

由于每个样本点对应的相似样本都是独立的, 可将式(7)拆分优化. 以 x_j 为例, 有

$$\begin{aligned} & \min_{S_{1j}} s_{1j}^T (\mu (I_{1*k} \times x_j - D^{x_j}) + \\ & \gamma (I_{1*k} \times p_j - D^{p_j})) s_{1j}, \end{aligned} \quad (8)$$

其中 D^{x_j} 和 D^{p_j} 分别表示 x_j 的相似邻居集合的特征矩阵和候选标记矩阵. 显然, 式(8)是一个凸二次规划问题, 可以利用相关工具求解.

对于 S_1 而言, 每次只优化与当前样本相关的 k 个变量, 当优化所有样本后就可以得到 S_1 .

其次, 固定 S_1 和 W , 优化 P . 固定 S_1 和 W 后, 有

$$\begin{aligned} & \min_P \sum_{i=1}^m \|L(x_i, W) - P\|_2^2 + \\ & \mu \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_1} s_{1ij} p_i \right\|_2^2 - \\ & \eta \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_2} s_{2ij} p_i \right\|_2^2; \end{aligned}$$

$$\text{s.t. } \forall 1 \leq i \leq m, 0 \leq P \leq Y, \sum_i P_{ji} = 1. \quad (9)$$

在式(9)中,前两项是凸的,而后一项是凹的,因此,它是一个带约束的凹凸问题. 本文利用传统的凹凸过程CCCP^[17]进行求解,根据凹凸过程,式(9)可被等价替换成 $C_1 - \widetilde{C}_2$,其中

$$C_1 = \sum_{i=1}^m \|L(x_i, W) - P\|_2^2 + \mu \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_1} s_{1_{ij}} p_i \right\|_2^2.$$

令 $p_i^{(k)}$ 表示第 k 次迭代时 P 的值, \widetilde{C}_2 可用一阶泰勒函数近似得到,即

$$\begin{aligned} \widetilde{C}_2 &= 2\eta \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_2} s_{2_{ij}} p_i^{(k)} \right\|_2^2 - \\ &\quad \eta \sum_{j=1}^m \left\| p_j^{(k)} - \sum_{(x_i, x_j) \in \xi_2} s_{2_{ij}} p_i^{(k)} \right\|_2^2. \end{aligned} \quad (10)$$

根据凹凸过程,式(9)可被等价替换成

$$\begin{aligned} \min_P \sum_{i=1}^m \|L(x_i, W) - P\|_2^2 + \\ \mu \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_1} s_{1_{ij}} p_i \right\|_2^2 - \\ 2\eta \sum_{j=1}^m \left\| p_j - \sum_{(x_i, x_j) \in \xi_2} s_{2_{ij}} p_i^{(k)} \right\|_2^2; \\ \text{s.t. } \forall 1 \leq i \leq m, 0 \leq P \leq Y, \sum_i P_{ji} = 1. \end{aligned} \quad (11)$$

将模型输出 $L(x_i, W)$ 用 O 表示, P 、 O 、 Y 向量化表示为 $\tilde{p} \in \mathbf{R}^m q$, $\tilde{o} \in \mathbf{R}^m q$, $\tilde{y} \in \mathbf{R}^m q$,式(11)等价于

$$\begin{aligned} \min_P \frac{1}{2} \tilde{p}^T H \tilde{p} + f^T \tilde{p}; \\ \text{s.t. } \forall 1 \leq i \leq m, 0 \leq P \leq Y, \sum_i P_{ji} = 1. \end{aligned} \quad (12)$$

其中: $H = I_{q \times q} \otimes \mathbf{T}_{m \times m} + (I_{q \times q} \otimes \mathbf{T}_{m \times m})^T$, $T = \mu(S_1 S_1^T + (S_1^T I_{m \times m} S) \odot I_{m \times m} - 2S_1^T) + I_{m \times m}$, $f = -\tilde{o} - \eta I_{m \times m} \otimes (I_{m \times m} - S_2) \tilde{p}^{(i)}$. 显然,式(12)是一个凸二次规划问题,可以利用相关工具进行求解. 最后,固定 S_1 和 P ,优化 W . 当固定 S_1 和 P 后,得到下式:

$$\min_W \sum_{i=1}^m \|L(x_i, W) - P\|_2^2 + \lambda \|W\|_2^2. \quad (13)$$

这就是一个简单的多元回归问题了,可以直接应用最小二乘法求解,有

$$W = (X^T X + \lambda I_{m \times m})^{-1} X^T P. \quad (14)$$

为了更好地拟合效果以适应非线性情况,可以利用核函数将特征空间向高维映射. 与大部分的消歧算法一样^[10-11],本文算法默认采用高斯核函数

$\kappa(x_i, x_j) = \exp(-\|x_i - x_j\|_2^2) / \sigma^2$, σ 表示训练样本之间的平均欧氏距离,则式(13)可表示为

$$\min_A \|KA - P\|_2^2 + \lambda \text{tr}(A^T K A), \quad (15)$$

其中 A 是将 W 映射到高维 $\phi(X)$ 的权重矩阵,令其导数为零,可得

$$A = (K + \lambda I_{m \times m})^{-1} P. \quad (16)$$

以上处理过程通过式(7)~(16)便可完成一次迭代,接着循环前面的步骤,直到算法收敛或者达到最大迭代次数为止,取模型输出最大值的候选标签作为最终结果,即

$$y^* = \arg \max_{j \in [q]} \sum_{i=1}^m a_{ij} \kappa(x, x_i). \quad (17)$$

PL-ODG的算法伪代码如算法1所示.

算法1 离异图消歧引导的偏标记学习 PL-ODG.

输入: D : 偏标记训练集 $\{(x_i, y_i) | 1 \leq i \leq m\}$; k : 相似样本数; 参数: $\lambda, \mu, \eta, \gamma, x^*$: 未知示例.

1) statement of your algorithm

2) 初始化置信度矩阵 P

3) 利用高斯核构造核矩阵 $K = [\kappa(x_i, x_j)]_{m \times m}$

4) repeat

5) 根据定义3和定义4,分别构造相似图 G_1 和离异图 G_2

6) 通过式(8)更新 G_1 的权重矩阵 S_1

7) 通过式(16)更新权重矩阵 A

8) 模型输出 $O = KA$

9) 通过式(12)更新置信度矩阵 P

10) until 算法收敛或达到最大迭代次数

输出: $y^* = \arg \max_{j \in [q]} \sum_{i=1}^m a_{ij} \kappa(x, x_i)$.

3 实验结果与分析

本文使用5个真实数据集,其中包括Lost^[18]、MSRCv2^[19]、FG-NET^[20]、Soccer Player^[21]、Yahoo! News^[22]. 比较了5种基于辨识的偏标签学习算法,有SURE^[23]、PL-AGGD、SDIM、PL-BLC^[13]、PRODEN^[14],以及两种基于平均的算法PLKNN、IPAL,每种算法都采用文献中推荐的超参数设置. 对于PL-ODG,采用的参数设置为 $\gamma, \eta, \mu, \lambda$,分别从0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, ..., 1中搜索, k 在10, 15, 20, 25, 30中搜索.

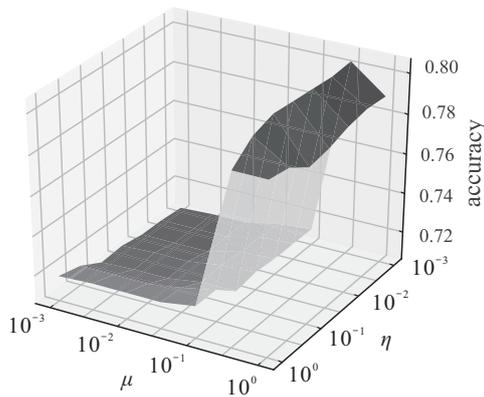
通过10倍交叉验证的分类精度的平均值和标注偏差来衡量算法的准确性,实验结果如表1所示,其中●表示PL-ODG在0.05显著性水平下的 t 检验优于其他算法.

表1 真实数据集上不同算法评估结果(平均精度±标准差)

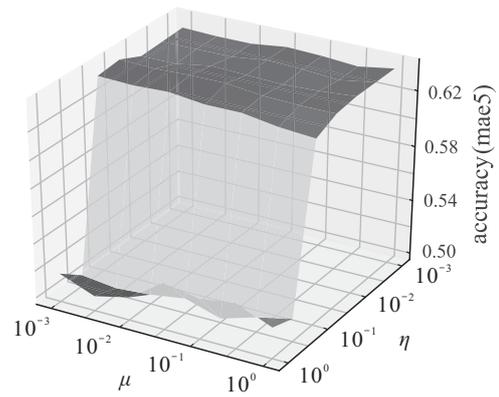
method	lost	MSRCv2	soccer player	Yahoo! news	FG-NET	FG-NET(MAP3)	FG-NET(MAP5)
PLKNN	0.434±0.052●	0.421±0.031●	0.494±0.023●	0.486±0.035●	0.037±0.023●	0.271±0.043●	0.455±0.024●
IPAL	0.679±0.052●	0.527±0.036●	0.539±0.013●	0.601±0.013●	0.052±0.013●	0.371±0.019●	0.551±0.025●
SURE	0.781±0.036●	0.486±0.042●	0.536±0.015●	0.645±0.025●	0.068±0.025●	0.377±0.310●	0.505±0.438●
PL-AGGD	0.747±0.046●	0.518±0.025●	0.517±0.007●	0.646±0.005●	0.0858±0.029	0.465±0.38●	0.619±0.534●
SDIM	0.804±0.053	0.516±0.042●	0.575±0.022	0.667±0.035	0.075±0.021●	0.454±0.378●	0.609±0.412●
PL-BLC	0.803±0.041	0.525±0.031	0.542±0.024●	0.649±0.041●	0.074±0.015●	0.426±0.392●	0.569±0.494●
PRODEN	0.811±0.023	0.456±0.033●	0.553±0.052●	0.674±0.019	0.049±0.031●	0.395±0.251●	0.550±0.507●
ours.	0.808±0.030	0.532±0.029	0.567±0.023	0.670±0.023	0.0878±0.025	0.483±0.391	0.634±0.544

表1中,由于FG-NET利用人脸预测人的年龄(1~100岁),其候选标签歧异度(候选标签的平均数量)较大,导致算法的分类精度较低.为了更好地评估这一任务,本文额外设置了两个实验,采用预测年龄与真实年龄差不超过3年(MAP3)、5年(MAP5),则视为分类正确.

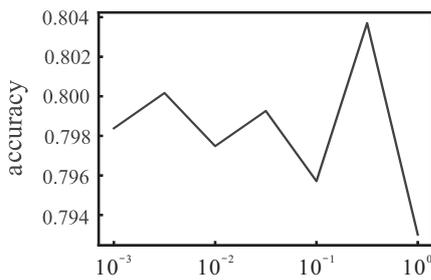
从表1中可以看出,本文算法在5个数据集上的表现不输于所对比的其他算法,且都达到了较好的效果.此外,在数据集歧异度较高的FG-NET数据集上的3个实验,本文算法几乎优于所有对比算法,这表明本文算法在处理歧异度更高的数据集时具有明显的优势.



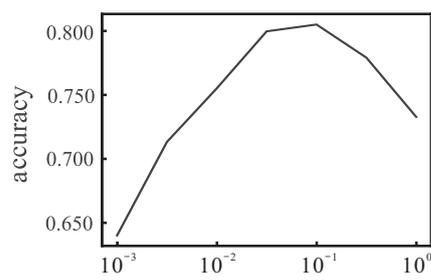
(a) 在Lost数据集上改变 μ, η



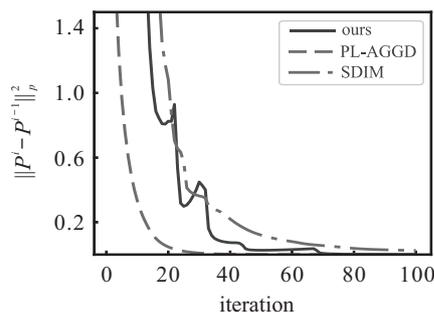
(b) 在FG-NET数据集上改变 μ, η



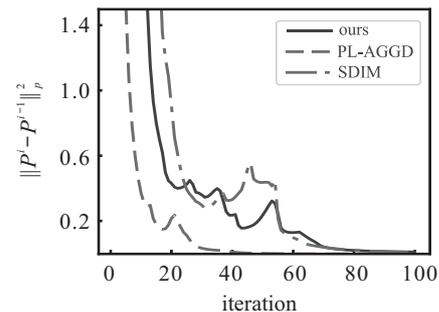
(c) 在Lost数据集上改变 λ



(d) 在Lost数据集上改变 γ



(e) 在Lost数据集上收敛轮数



(f) 在MSRCv2数据集上收敛轮数

图3 算法收敛性分析

本文对 PL-ODG 算法的收敛性进行了分析, 针对算法中除 k 值以外的 4 个超参数, 通过控制变量的方式, 对每个参数的敏感性进行分析. 在 4 个超参数中, μ 和 η 是对训练精度影响最大的两个参数, 它们分别用于控制相似图和离异图的权重系数. 图 3(a) 和图 3(b) 是在 Lost 和 FG-NET 数据集上改变两个参数后, Lost 的分类精度以及 FG-NET 的分类误差小于 5 的精度变换情况. 可以发现, 当 μ 和 η 小于 10^{-3} 时对于精度的影响不大. 另外, 在歧异度低的数据集 Lost 上, η 的作用更为明显, 当 η 大于 0.1 后, 分类精度显著上升. 在歧异度高的数据集 FG-NET 上, μ 的作用较为明显, 当 μ 小于 0.5 后, 分类精度大幅提升. λ 用于优化相似图时调整样本标签和样本的权重, 图 3(c) 是对其在 Lost 上的变换影响进行的测试结果, 该值的变化对于算法整体准确率的影响波动较小, 可以在需要精确调优时进行调整. 最后, 图 3(d) 是对 γ 在 Lost 的变换影响的测试结果, γ 用于控制 L_2 正则的权重以防止模型过拟合. 可以看到, 当 γ 过小 (小于 0.05) 或者过大 (大于 0.5) 时, 模型会因为过拟合或者欠拟合而导致分类精度显著下降. 因此, 在实际使用中本文建议设置 $k = 20, \lambda = 0.1, \mu = 0.3, \eta = 0.3, \gamma = 0.05$, 以达到较为稳定的效果. 在此基础上, 结合数据集的歧异度调整超参数: 歧异度高时, 调低 μ 和 η 系数, 增加 μ 和 η 的比值; 歧异度低时, 调高 μ 和 η 系数, 增加 μ 和 η 的比值.

图 3(e) 和图 3(f) 给出了在 Lost 和 MSRCv2 两个数据集上与同类型消歧算法进行算法收敛性的对比结果. 其中, 纵坐标表示置信度矩阵每轮迭代的 L_2 正则, 用于度量每次迭代后置信度矩阵的差异. 可以发现: 随着迭代次数增多, 置信度矩阵逐渐开始收敛; PL-ODG 算法在两个数据集上均经历 70 次左右迭代后才收敛, 其收敛速度比 PL-AGGD 慢, 比 SDIM 略快.

4 结 论

本文研究了如何利用先验知识引导偏标记学习消歧的问题, 通过分析样本离异点的特点, 提出了一种离异图引导的偏标记学习方法 PL-ODG. 文中提出的方法有效利用了标签空间的先验知识, 改善了样本离异点对特征相似图所带来的消极影响, 并合理利用离异点信息指导算法的消歧. 在真实数据集上的实验表明, 本文提出的算法能够达到较好的消歧水平.

此外, 当前利用嵌入样本先验知识进行消歧的方法都是基于全局信息的, 且采用了 EM 算法进行迭

代, 它们需要在每次训练前遍历全量训练样本, 因此, 较之于随机优化的算法而言在大数据集上难以保障高效性. 但是, 嵌入先验知识进行消歧的思想又可以极大程度提高训练样本的价值, 达到更好的消歧效果. 未来期望进一步结合两种算法的特点, 在随机优化过程中嵌入先验知识进行消歧, 以适应大规模数据的处理.

参考文献 (References)

- [1] Hüllermeier E, Beringer J. Learning from ambiguously labeled examples[J]. Intelligent Data Analysis, 2006, 10(5): 419-439.
- [2] Cour T, Sapp B, Taskar B. Learning from partial labels[J]. The Journal of Machine Learning Research, 2011, 12(4): 1501-1536.
- [3] Chen C H, Patel V M, Chellappa R. Learning from ambiguously labeled face images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(7): 1653-1667.
- [4] Liu L, Dietterich T. A conditional multinomial mixture model for superset label learning[J]. Advances in Neural Information Processing Systems, 2012, 25: 548-556.
- [5] Liu L P, Dietterich T G. Learnability of the superset label learning problem[C]. Proceedings of the International Conference on Machine Learning. Beijing, 2014: 1629-1637.
- [6] Gong C, Liu T L, Tang Y Y, et al. A regularization approach for instance-based superset label learning[J]. IEEE Transactions on Cybernetics, 2018, 48(3): 967-978.
- [7] Luo J, Orabona F. Learning from candidate labeling sets[J]. Advances in Neural Information Processing Systems, 2010, 23: 1504-1512.
- [8] Zhou D Y, Zhang Z K, Zhang M L, et al. Weakly supervised POS tagging without disambiguation[J]. ACM Transactions on Asian and Low-Resource Language Information Processing, 2018, 17(4): 1-19.
- [9] Zhang M L, Zhou B B, Liu X Y. Partial label learning via feature-aware disambiguation[C]. KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, 2016: 1335-1344.
- [10] Wang D B, Li L, Zhang M L. Adaptive graph guided disambiguation for partial label learning[C]. KDD'19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. San Francisco, 2019: 83-91.
- [11] Feng L, An B. Partial label learning by semantic difference maximization[C]. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, 2019: 2294-2300.
- [12] 陈鸿昶, 谢天, 高超, 等. 候选标记信息感知的偏标记学习算法[J]. 电子与信息学报, 2019, 41(10):

- 2516-2524.
(Chen H C, Xie T, Gao C, et al. Candidate label-aware partial label learning algorithm[J]. Journal of Electronics & Information Technology, 2019, 41(10): 2516-2524.)
- [13] Yan Y, Guo Y H. Partial label learning with batch label correction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 6575-6582.
- [14] Lv J, Xu M, Feng L, et al. Progressive identification of true labels for partial-label learning[C]. International Conference on Machine Learning. Vienna, 2020: 6500-6510.
- [15] Cabannes V, Rudi A, Bach F. Structured prediction with partial labelling through the infimum loss[C]. International Conference on Machine Learning. Vienna, 2020: 1230-1239.
- [16] Wen H, Hang H, Liu J, et al. Leveraged weighted loss for partial label learning[C]. Proceedings of 36th International Conference on Machine Learning. Vienna, 2021: 11091-11100.
- [17] Yuille A L, Rangarajan A. The concave-convex procedure[J]. Neural Computation, 2003, 15(4): 915-936.
- [18] Cour T, Sapp B, Jordan C, et al. Learning from ambiguously labeled images[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, 2009: 919-926.
- [19] Liu L, Dietterich T G. A conditional multinomial mixture model for superset label learning[C]. Proceedings of The Advances in Neural Information Processing Systems. Nevada, 2012: 548-556.
- [20] Panis G, Lanitis A, Tsapatsoulis N, et al. Overview of research on facial ageing using the FG-NET ageing database[J]. IET Biometrics, 2016, 5(2): 37-46.
- [21] Zeng Z N, Xiao S J, Jia K, et al. Learning by associating ambiguously labeled images[C]. 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, 2013: 708-715.
- [22] Guillaumin M, Verbeek J, Schmid C. Multiple instance metric learning from automatically labeled bags of faces[C]. Computer Vision—ECCV 2010. Berlin: Springer, 2010: 634-647.
- [23] Feng L, An B. Partial label learning with self-guided retraining[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 3542-3549.

作者简介

胡峰(1978—), 男, 教授, 博士, 从事粗糙集、机器学习、数据挖掘等研究, E-mail: hufeng@cqupt.edu.cn;

刘鑫(1996—), 男, 硕士生, 从事数据挖掘、自然语言处理的研究, E-mail: 434677724@qq.com;

邓维斌(1978—), 男, 教授, 博士, 从事智能决策与知识发现、智能信息处理等研究, E-mail: dengwb@cqupt.edu.cn;

代劲(1978—), 男, 教授, 博士, 从事粗糙集、云模型、数据挖掘等研究, E-mail: daijing@cqupt.edu.cn;

刘群(1969—), 女, 教授, 博士生导师, 从事复杂网络、非线性理论、神经网络、智能信息处理等研究, E-mail: liuqun@cqupt.edu.cn.

下 期 要 目

- 进化超多目标优化研究进展及展望..... 肖人彬, 等
- 多目标动态优化中Pareto随机合作博弈研究综述..... 张维海, 等
- 多任务优化算法综述..... 程美英, 等
- “双碳”背景下联合配送冷链物流模型及其求解算法..... 陈雨蝶, 等
- 数据驱动选择策略的多目标差分进化算法..... 侯莹, 等
- 基于残差注意网络的端到端手写文本识别方法..... 王寅同, 等
- 线性时序逻辑引导的安全强化学习..... 李保罗, 等
- 基于残差U型卷积网络的海面风速多步时空预测..... 谢莒芃, 等
- 考虑元素烧损和仓储成本的特种铝合金配料优化..... 张浩, 等
- 融合双目视觉和2D激光雷达的室外定位..... 刘净轩, 等
- 基于功能性电刺激的腕部震颤抑制系统线性自抗扰控制..... 刘艳红, 等
- 基于AC-YOLO的路面落叶检测方法..... 缪燕子, 等
- 基于准最大最小模型预测控制的AUV视觉对接..... 岳伟, 等